PATTERNS OF SURVEY ATTRITION AND RELUCTANT RESPONSE IN THE 1996 MEPS

Steven B. Cohen, Steven R. Machlin and Jim Branscome, Agency for Health Care Policy and Research
Steven B. Cohen, AHCPR, 2101 E. Jefferson Street, Suite 500, Rockville, Maryland 20852

1. Introduction

In national household health care surveys, a significant amount of survey resources are often allocated to obtain the participation of households that constitute the last five to ten percent of the overall survey response rate. A substantial number of households that respond toward the end of the survey field period are characterized by an initial refusal to participate. In these household surveys, an overall survey response rate is specified as a required target, with the implicit assumption that failure to achieve the desired rate will have a deleterious impact on the accuracy and precision of resultant estimates. When the response rate targets are in jeopardy of not being met, the data collection strategy employed makes concerted use of nonresponse conversion techniques in order to improve the overall survey response rate. Applications of these nonresponse conversion techniques are not cost neutral, and can result in significant increments to data collection costs.

In panel designs with multiple waves of data collection, the overall survey response rate is a multiplicative function of the wave specific response rates. The 1996 Medical Expenditure Panel Survey (MEPS) follows this model, requiring five rounds of data collection with the same panel of sampled households, to acquire data on health care use, expenditures, insurance coverage and sources of payment that cover two consecutive calendar years. Gaining an understanding of the factors that distinguish the cooperative respondents, the survey participants that require use of nonresponse conversion techniques to maintain their cooperation (reluctant respondents), and the initial participants that eventually drop out of the survey (part-year respondents) is essential from both an estimation and data collection perspective. To inform the specification of nonresponse adjustment strategies in MEPS to correct for survey attrition, this study attempts to identify the characteristics that distinguish survey participants across panels from the part-year respondents. In addition, the study includes an analysis to identify those factors that distinguish the cooperative respondents, the reluctant respondents and the part-year respondents, to better inform the MEPS data collection effort. The investigation also examines the implications of a data collection strategy that would not convert initial survey refusals by studying the effect on resultant survey estimates and their precision that is attributable to the exclusion of reluctant survey respondents. The results provide insights regarding the accuracy of MEPS survey estimates that characterize the second year of the longitudinal panel, where response rates are comparable to those simulated by the exclusion of reluctant respondents.

2. Background

The level of bias in survey estimates attributable to nonresponse in household surveys can be expressed as a multiplicative function of the level of nonresponse experienced by the survey and the magnitude of the difference in the parameter estimates that distinguish respondents from nonrespondents (Madow, Nisselson, and Olkin, 1983; Cochran, 1977). Consequently, the achievement of a high rate of response for a survey will serve to limit the impact of nonresponse bias on resultant survey estimates. Repeated attempts to gain the cooperation of eligible respondents in the form of callbacks are an established component of most household survey data collection plans. The primary objective of this nonresponse conversion technique is to enhance the overall survey response rate and to achieve a reduction in nonresponse bias. With few exceptions (Berk, 1984; Berk, Wilensky and Cohen, 1984), the subset of a targeted sample that respond on a first survey contact differ from the overall sample (Madow, Nisselson, and Olkin, 1983; Jessen, 1978). To the extent that one can determine the factors associated with distinguishing the respondents from the nonrespondents, and incorporate this information into the development of the nonresponse adjustments to the survey specific estimation weights, a reduction in nonresponse bias is attainable.

It has also been noted that survey respondents that initially refuse differ from cooperative respondents on demographic characteristics, which often translate to significant differences in the core analytic survey variables (Fitzgerald and Fuller, 1982; O' Neil, 1979). Although the desirability of enhancing the survey response rate through the implementation of callback interviews is not in dispute, the associated data collection cost places a significant constraint on their application. With a fixed survey budget, it is essential to insure judicious utilization of the callback interview techniques. A desirable approach would be to determine in advance of the survey the optimal number and allocation of callback interviews that minimize the mean square error of survey estimates for fixed cost.

3. The 1996 MEPS Household Component

The Medical Expenditure Panel Survey (MEPS) household component is an ongoing annual survey, with each sample panel collecting data over a 30 month period, with five in-person CAPI administered interviews, to obtain information that covers two consecutive calendar

years. The MEPS collects data on the specific health services that Americans use, how frequently they use them, the cost of these services and how they are paid, as well as data on the cost, scope, and breadth of private health insurance held by and available to the U.S. population. MEPS is unparalleled for the degree of detail in its data, as well as its ability to link health service medical expenditures and health insurance data to the demographic, employment, economic, health status, utilization of health services, and other characteristics of survey respondents. Moreover, MEPS is the only federally sponsored national survey that provides a foundation for estimating the impact of changes in sources of payment and insurance coverage on different economic groups or special populations of interest, such as the poor, elderly families, veterans, the uninsured, and racial and ethnic minorities. The survey is sponsored by the Agency for Health Care Policy and Research with co-sponsorship from the National Center for Health Statistics. Westat and the National Opinion Research Center (NORC) are the data collection organizations for the 1996 MEPS Household Survey.

4. MEPS Round 1 Field Results

The MEPS sample is a nationally representative sub-sample of households that were part of the 1995 NHIS. Consequently, the response rate that has implications in the development of national estimates from MEPS is a function of the response rates to both surveys. More specifically, the overall Round 1 MEPS response rate of 77.7 percent is derived as the product of the following three component rates: (1)the NHIS response rate achieved for the households eligible for the MEPS (93.9 percent);(2)the proportion of NHIS units selected that had sufficient information to permit MEPS data collection (99.6 percent);(3)the MEPS Round 1 reporting unit response rate (83.1 percent).

5. Characteristics of MEPS Nonrespondents - Round 1

Prior MEPS findings revealed that the nonresponding dwelling units were more likely to consist of single or two person households and had a higher representation in large metropolitan statistical areas (Cohen and Machlin, 1997). These results are generally consistent with the profiles that distinguish survey response status in prior cycles of national medical expenditure surveys. Single person households and households located in large metropolitan areas are generally more difficult to schedule interviews with, posing additional challenges to a field staff in obtaining their survey participation. In addition, the comparison with respect to family income revealed that the nonresponding households had a significantly higher representation of high income classifications. Furthermore, the comparison of the health specific summary measure, which classified dwelling units into distinct categories based on the ability of their

members to perform personal care activities or other routine needs and by age, demonstrated that nonresponding households were more likely to include healthy elderly members. The nonresponding households were dramatically less likely to have responded to the income question, provided their telephone numbers, or responded to the work classification question in the 1995 NHIS.

6. MEPS Round 1 Respondents Distinguished by Level of Cooperation: Round 1-2

Table 1 below identifies the level of cooperation received during the first two rounds of the MEPS survey from those who responded in round 1. Those cases that required a follow-up interview to convince a reluctant respondent to report are referred to in the table as Initial Refusals. In subsequent sections of this paper, we will be assessing the characteristics of these reluctant respondents and how they compare to and contrast with cooperative respondents and non-respondents.

7. Characteristics of the MEPS Nonrespondents in Round 2

Within the 9,488 reporting units that responded to the first round of interviewing in the 1996 MEPS, there were 23,612 survey participants who were members of the civilian non-institutionalized population. These respondents served as the underlying sample from which the Round 1 1996 MEPS person level national population estimates were derived. Of the subset of 23,540 MEPS Round 1 survey participants also eligible for Round 2, 1,393 individuals, or 5.9 percent did not respond in the MEPS Round 2 interview (Table 1). While the conditional Round 2 person level response rate was 94.1 percent for the overall eligible Round 1 sample, the Round 2 response rate for the subset of individuals (1,710 or 7.3 percent) who resided in households that initially refused to participate in the first round of MEPS was only 86.9 percent, indicating a clear association between the probability of future survey participation and prior indications of resistance to respond.

To help ascertain the potential level of nonresponse bias that was attributable to survey attrition in Round 2 of the MEPS, it was necessary to determine whether the individuals in households that did not continue their participation in the MEPS differed systematically from their Round 2 respondent counterparts. To facilitate these comparisons, the following demographic and socio-economic profiles that characterized these two distinct respondent groups at the time of the first MEPS interview were compared: Age (<20, 20-29, 30-44, 45-64, 65+);Race/ethnicity (Hispanic, Black/Non-Hispanic, Other); Sex; Size of MEPS Reporting Unit (1,2,3,4,5+);MSA class (MSA, nonMSA);Region (northeast, midwest, south, west);Marital Status (married, widowed, divorced, separated, never married);Education

(under age 5, grades 0-8, grades 9-11, grade 12, 1-3 years of college, and 4+years of college);Employment (currently employed, unemployed or <16 of age); Health Insurance (private-employer sponsored, private-nonemployer sponsored, public, uninsured);Health Status (excellent, very good, good, fair, poor); Round 1 Interview Classification (no initial refusal, initial refusal)

In order to ascertain the primary factors that differentiated the MEPS Round 2 respondents from their nonresponding counterparts, a weighted logistic regression analysis was implemented. The analysis also adjusted the estimated variances of model coefficients for survey design complexities (Shah et al., 1996). This multivariate analysis allowed for a determination of the dominant factors associated with survey attrition, after controlling for the demographic, socio-economic and health related measures under study, in addition to a variable that indicated whether the household initially refused to participate in the first round of MEPS data collection. In the logistic regression, MEPS Round 2 nonrespondents were identified by a value of Y=1, with the remaining Round 2 respondents identified by the value of Y=0. A backwards elimination model specification technique was then implemented in order to identify the core set of factors that were determined to be significant predictors of MEPS survey attrition at the alpha = .05 level.

Table 2 provides a summary of the core set of factors that were determined to be significant predictors of MEPS Round 2 survey attrition, based on the results of the logistic regression analysis under consideration. The results of these analyses revealed that individuals in households that initially refused to participate in the first round of the MEPS were significantly more likely to not participate in Round 2, relative to individuals in households that were less resistant to respond. As in the first MEPS interview, individuals who resided in metropolitan statistical areas were also more likely to end their survey participation than those Round 1 respondents residing in non-metropolitan areas. It was also noted that individuals residing in reporting units of size 1 or 2 had a higher likelihood of maintaining their participation in MEPS relative to individuals in the largest reporting units. This association may be suggestive of a significant detrimental impact of survey burden, as measured by length of interview, on continued survey participation in a longitudinal data collection effort.

Age was also determined to be significantly associated with survey attrition. Individuals aged 65 and older had a higher likelihood of dropping out of the survey relative to children. In addition, individuals who were married or separated had a higher likelihood of maintaining their participation in MEPS relative to individuals who were never married. Controlling for these significant factors,

the race/ethnicity, sex, region of residence, education level, employment status, health insurance coverage and health status did not significantly differentiate MEPS Round 2 survey response status.

8. MEPS Full Year Nonresponse Adjustment

Since survey nonresponse is potentially a significant source of error or bias in survey estimates, the MEPS full year sampling weights included an adjustment for survey attrition to help reduce potential the impact of bias from survey nonresponse. In the MEPS, a weighting class nonresponse adjustment was implemented, which assumes that nonresponding sampling units would have responded in a similar manner as respondents with similar socio-demographic and economic characteristics within the same adjustment class. Properly designed, a weighting class nonresponse adjustment strategy will result in a reduction in nonresponse bias. The technique requires a partitioning of the sample into mutually exclusive classes, with classification information available for both responding and nonresponding units (Cox and Cohen, 1985).

Consequently, the person level estimation weight adjusted for MEPS full year 1996 nonresponse, $W(i)$ takes form: $W(i) = A(c) \times W1(i)$,

where: $W1(i)$ is the MEPS Round 1 nonresponse adjusted weight for the ith MEPS sample respondent, which reflects the reciprocal of the person's selection probability for MEPS and a poststratification adjustment to March 1996 CPS population totals, and $A(c)$ is the nonresponse adjustment for class c.

The results of the multivariate analyses of the characteristics that distinguish MEPS Round 2 respondents from their nonrespondent counterparts were used to identify the most important measures to include in the specification of a nonresponse adjustment to the MEPS sampling weights to correct for potential nonresponse bias at the person level. This effort helped facilitate the development of estimation weights for the 1996 MEPS Full Year Household Survey Health Care Utilization data release according to schedule requirements, without incurring additional contractor costs for the conduct of a comparable nonresponse profiling analysis. The cells that defined the nonresponse adjustments were based on cross-classifications of the measures identified in the logistic regression analyses as significant predictors of survey attrition.

Ideally, an analysis that distinguished part-year respondents from the full year MEPS participants, which incorporated information from Round 3 of the survey, would have been the best approach to consider to inform the specification of the full year nonresponse adjustment. However, the additional Round 3 data covering 1996 was unavailable at the time of this analysis and the conditional MEPS Round 3 response rate was considerably higher

than Round 2 experience (96 percent). There were also several advantages, from both a survey cost and schedule perspective, that would be achieved by using the Round 2 study results of survey attrition patterns to inform the nonresponse adjustment strategy. Furthermore, additional improvements to the accuracy of survey estimates would be realized by inclusion of a poststratification adjustment to the final MEPS person level estimation weight, based on December 1996 CPS population totals cross-classified by age, race/ethnicity and gender. Once the full year MEPS data files are available, an expansive analysis of survey attrition that includes Round 3 data will be conducted. This planned study will offer an opportunity to re-examine the weighting class adjustments that were identified through the preliminary analysis, and implement revisions if necessary. A strong convergence in analytical findings will provide confirmation that no revisions to the weighting classes that defined the person level nonresponse adjustments are necessary.

9. Demographic and Socio-Economic Characteristics of the Reluctant MEPS Respondents: Round 1 and 2

Of the 22,147 individuals with MEPS responses to both the Round 1 and Round 2 interviews, 2,109 survey participants or 9.5 percent resided in households that initially refused to participate in one of these two core rounds of data collection (Table 1). A failure to convert these initial survey refusals would have resulted in an overall MEPS Round 2 person level response rate of only 66.1 percent (.777 x .8512), which was significantly lower than the overall Round 2 person level response rate of 73.1 percent (.777 x .9408) that was achieved for the survey. Consequently, the data collection strategy employed in MEPS, that emphasized the application of nonresponse conversion techniques, was successful in improving resultant survey response rates. By virtue of their hesitancy to participate in the MEPS, it was hypothesized that they differed from their cooperative counterparts on a number of dimensions, and their successful inclusion in the MEPS as survey participants would have a beneficial effect in reducing the impact of nonresponse bias on resultant MEPS survey estimates.

To determine whether the underlying assumption that individuals residing in participating households that initially refused survey participation significantly differed from their more cooperative counterparts on a set of demographic, socio-economic and health related measures, a weighted logistic regression analysis was also implemented. In the logistic regression, the responding individuals residing in households that initially refused survey participation in either of the first two rounds of data collection were identified by a value of Y=1 and referred to as reluctant respondents, with the remaining cooperative respondents identified by the value of Y=0. All of the demographic, socio-economic and health

related measures considered in the analysis of attributes of nonresponse were also considered in this exploratory analysis.

Table 3 provides a summary of the core set of factors that were determined to be significant predictors of the reluctant respondents in MEPS, based on the results of the logistic regression analysis under consideration. As in the 1987 National Medical Expenditure Survey (NMES), MEPS survey participants who were located in metropolitan statistical areas were also more likely to reside in households that initially refused survey participation, relative to the MEPS respondents residing in non-metropolitan areas (Cohen and Carlson, 1995). Again, similar to the pattern identified in the 1987 NMES, MEPS survey respondents located in the Northeast and West were significantly more likely to be reluctant respondents, relative to the survey participants from the Southern region. It was also noted that individuals aged 65 and older were substantially more likely to be in responding households that initially refused to participate in MEPS, relative to all the other age groups. Education level was also determined to be significantly associated with reluctance to participate in the survey. Individuals with four or more years of college education had a lower likelihood of residing in households that initially refused to participate in MEPS, relative all other individuals. In addition, individuals who were either unemployed or not in the labor force unemployed were significantly more likely to be reluctant respondents, relative to the survey participants who were employed. Controlling for these significant factors, the race/ethnicity, sex, marital status, reporting unit size, health insurance coverage and health status did not significantly differentiate MEPS respondents with respect to an initial refusal to participate.

This analysis included all MEPS Round 2 respondents, in order to provide a better understanding of the implications on survey estimates if the households that initially refused to participate were not converted to survey participants in either Rounds 1 or 2 of the MEPS. To better understand the dynamics of the MEPS field effort, and to identify the profiles of the decision makers with respect to survey participation, the analysis was replicated for the restricted sample of individuals that were directly interviewed. Although the effects of education level and employment status were not significant factors in distinguishing the reluctant respondents from their cooperative counterparts, the remaining factors of age, MSA and regional location retained their significance as noted for the full sample. The results suggest that MEPS households located in MSAs and in the Northeastern and Western regions of the country, and those potential survey respondents aged 65 and older, would experience lower survey response rates

in the absence of targeted nonresponse conversion techniques.

10. Respondents in Households that Initially Refused to Participate in MEPS: More Like Part-Year Respondents or Cooperative Respondents?

Overall, the respondents in households that initially refused to participate in the MEPS differed from cooperative respondents on several of the same characteristics that also distinguished MEPS nonrespondents from the respondents. These characteristics included age and location of residence with respect to MSA status. In order to more directly determine whether the respondents in households that initially refused to participate in the MEPS were more similar to the part-year respondents or the cooperative respondents, a multinomial logistic regression analysis was conducted. All of the same demographic, socio-economic and health related measures considered in the prior logistic regression analyses were also considered in this more expansive investigation, which included all Round 1 respondents eligible for data collection in Round 2.

Table 4 provides a summary of the core set of factors that were determined to be significant factors that distinguished the MEPS respondent groups under consideration, based on the results of the multinomial logistic regression analysis under consideration. The multivariate results revealed that age, residence in a metropolitan area, and employment status were significant factors in distinguishing the alternative respondent groups. More specifically, elderly (65+), individuals residing in MSAs, and individuals who were not employed were significantly less likely to cooperate without any resistance in the MEPS relative to the younger survey participants, individuals located in non-metropolitan areas and the employed. Alternatively, age, residence in a metropolitan area and employment status did not significantly differentiate the likelihood of survey attrition and reluctant response, indicating similar profiles on these dimensions for the two resistant respondent groups. As in the union of the more constrained logistic regression analyses, the measures of reporting unit size, region, marital status and education level were determined to be significant in differentiating the likelihood of classification in one of the three respondent groups under investigation. Individuals reporting units of size 1 or 2 had a higher likelihood of maintaining their participation in MEPS, as residing in either cooperative or reluctant respondents, relative to individuals in the largest reporting units. For region, residence in the west was associated with a higher probability of a reluctant respondent classification relative to residence in the south. Individuals who were married or separated had a higher likelihood of participating in MEPS either as a

cooperative or reluctant respondents, relative to individuals who were never married. In addition, individuals with four or more years of college education were the least likely to reside in households that initially refused to participate in MEPS.

Based on the results of this analysis, the profiles that are most highly correlated with a classification of reluctant respondent are not coincident with those that distinguish either the cooperative respondents or the part-year respondents. Consequently, the findings indicate that the respondents in households that initially refused to participate in either of the first two rounds of data collection in MEPS are a distinct group. While more similar to the part-year respondents with respect to their profiles on age, MSA residence and employment status, the reluctant respondents more closely mirror the profiles for cooperative respondents on the dimensions of marital status and reporting unit size. Furthermore, the reluctant respondents are uniquely characterized by distinct profiles with respect to their level of education and their location by region.

11. Treatment of Reluctant Respondents as Nonrespondents: Impact on Survey Estimates

As described previously, a failure to interview respondents who were initially reluctant to cooperate in round 1 and/or round 2 would have reduced the overall 1996 MEPS person level response rate (prior to round 3) from 73.1 to 66.1 percent. To study the potential effect that this reduction would have had on survey estimates, the national estimates of access to care derived from the MEPS (Table 5) were re-computed after excluding the respondents who were reluctant to participate in round 1 and/or round 2 (Weinick and Zuvekas, 1997).

As shown in Table 5.1 below, most of the differences between the original estimates and the re-computed estimates of access to care derived from the MEPS that excluded reluctant respondents were negligible. Of the 100 estimates presented in Table 5, 83 changed by half a percentage point or less and only 2 changed by 1 percentage point or more, all nonsignificant differences (α =.05).The effect of excluding reluctant respondents on the precision of the estimates was also modest. Among the 100 estimates, the average standard error rose from 0.77 to 0.82, a relative increase of approximately 6.5 percent, and the maximum standard error increased slightly from 2.94 to 3.20.

To illustrate the effect of a more substantial reduction in cases on survey estimates, we also conducted an analysis based on national estimates of health insurance coverage derived from the MEPS (Table 6; Vistnes and Monheit, 1997). These estimates were based on round 1 respondents when the overall survey response rate was about 78 percent. The estimates in the table were re-computed after excluding round 2 non-respondents in

addition to respondents who were reluctant to participate in either round 1 or 2, which resulted in a simulated response rate of about 66.1 percent. The MEPS estimation weights were further adjusted for the exclusion of round 2 non-respondents and the respondents who were reluctant to participate.

As shown in the Table 6.1 below, about 90 percent of the differences between the original estimates and the re-computed estimates excluding reluctant respondents were less than 1 percentage point and 99 percent of the differences were less than 2 percentage points, all nonsignificant differences (α =.05). The combined effect on the precision of the estimates of excluding nonrespondents to round 2 and reluctant respondents to round 1 or round 2 was also modest. Among the 156 estimates, the average standard error increased from 1.07 to 1.13, a relative increase of approximately 6 percent. However, the maximum standard error had a somewhat more substantial increase from 3.51 to 4.17.

12. Summary

Within the context of the MEPS household component survey, we have examined and compared respondents who cooperatively participated in the survey, those who did not respond, and those who eventually responded after initially being reluctant to do so. Our analytical task was to identify characteristics of these groups and to note their similarities and differences. Our goals were to use this information to make weighting class non-response adjustments and to examine how this information might be used to improve future data collection strategies.

Reluctant respondents in the first round of the survey were significantly more likely to become non-respondents in the second round. These round two non-respondents were more likely to be in large households, located in large metropolitan areas, and were more likely to include elderly members. We were able to specify weighting class non-response adjustments using the Round 2 study results. Once the full year MEPS data files are available, that additional information will be used to confirm the weighting classes that have been developed and, if necessary, adjustments will be made to the weighting strategy in future years.

Reluctant respondents as a whole appear to be a distinctly separate group, sharing one set of characteristics with the cooperative respondent group, another set with those who refused during the second round of the survey, and yet a third set of characteristics that are uniquely their own. If no effort had been made to convert reluctant participants, the precision of our survey estimates would have dropped, but not substantially. Most of the simulated estimates changed by 1 percent or less and the standard errors of the estimates increased on a relative scale by approximately 6 percent. The results suggest a good level of accuracy can be expected from the MEPS survey estimates that characterize the second year of the longitudinal panel, where response rates are comparable to those simulated by the exclusion of reluctant respondents.

Regarding efforts to convert reluctant respondents; it is always beneficial to keep survey response rates as high as possible. This is particularly important for a panel survey with five rounds of data collection that cover two calendar years. Since the associated medical provider survey in MEPS requires the permission from household respondents to contact their medical providers, the achievement of a respectable response rate for this follow-back survey is dependent on the maintenance of a high response rate for the household survey. However, there is undeniably a cost associated with doing so. This may manifest itself in the form of cost overruns or the need to curtail other survey operations. The operational issue is to determine the appropriate level of effort to convert refusals. Furthermore, should that level of effort vary in intensity from round to round in a panel survey? Given the findings of our paper, a judicious use of refusal conversion will continue because it is beneficial to do so. To the extent that the MEPS data collection organizations are able to continue to improve upon the targeting of their refusal conversion efforts, making use of information to help determine where they are most likely to be successful, future savings in data collection costs may be achieved without loss in resultant survey response rates.